

Nakijken en nog eens

De correctie van het eindexamen geschiedenis

Ieder jaar in mei worden in Nederland de centrale schriftelijke eindexamens afgenomen.

Examenkandidaten in het voortgezet onderwijs maken opgaven die zijn geproduceerd door Cito onder verantwoordelijkheid van het College van Examens (CvE). Al dat examenwerk moet worden nagekeken door de eigen docent en gecontroleerd door een tweede corrector. Hoe gaat dat in zijn werk? In een groot onderzoek van Cito naar de praktijk van de eerste en tweede correctie wordt dat in een viertal publicaties beschreven. Dit artikel gaat vooral over het onderzoek naar de correctie van de geschiedenisexamens.

Stefan Boom is toetsdeskundige geschiedenis bij Cito. De auteur dankt Hans Kuhlemeier voor zijn bijdrage aan dit artikel.

Het gaat bij de centrale examens niet om kleine aantallen. Er werden in 2012 bijvoorbeeld in totaal ongeveer vijfhonderd examens ontwikkeld voor circa 200.000 kandidaten. Ongeveer 103.000 voor vmbo, 57.000 voor havo en omstreeks 40.000 voor vwo. De geschiedenisexamens voor havo en vwo maken daarvan een substantieel deel uit, met 39.000 havokandidaten (zeventig procent van het totaal aantal havokandidaten) en 20.350 vwo-kandidaten (circa vijftig procent van het totaal aantal vwo-kandidaten).¹ Deze aantallen namen sterk toe nadat geschiedenis een verplicht vak werd in de keuzeprofielen Cultuur en Maatschappij en Economie en Maatschappij. Vóór 2001, toen de vakkenkeuze nog niet aan verplichte profielen gebonden was, koos ongeveer een derde van de havisten en 45 procent van de vwo'ers voor geschiedenis als examenvak.

Dat eindexamen vormt de afsluiting van de middelbare school, en het behaalde diploma is een algemeen geaccepteerd bewijs dat die opleiding met 'goed gevolg' is voltooid. Daarvoor is het noodzakelijk dat kwaliteit, validiteit en aanvaardbaarheid van het examen gewaarborgd is. Een eerlijke beoordeling, die niet afhankelijk is van één docent, hoort daarbij. Daardoor kan het diploma van de middelbare school de civiele waarde behouden die het heeft: een toegangsbewijs tot een aansluitende vervolgopleiding, een kwaliteitsmerk bij sollicitaties.

Derde correctie

De noodzaak tot een goede monitoring van de correctie van de examens wordt al langer gevoeld. Het eerste onderzoek dat hier

besproken wordt, is een onderzoek naar de praktijk van de eerste en tweede correctie, het O2C-onderzoek. Bij een aantal vakken is een derde correctie uitgevoerd waarbij leerlingewerk opnieuw integraal is nagekeken. De derde correctoren hadden de opdracht de examenwerken zorgvuldig en objectief na te kijken volgens het correctievoorschrift. Aan dit onderzoek deden verschillende vakken mee, waaronder havo geschiedenis.² De uitkomsten van dit onderzoek waren opmerkelijk. Door de derde correctie gingen alle resultaten omlaag. Er was slechts een klein verschil tussen de eerste en tweede correctie (ook bij geschiedenis), maar de derde correctie maakte wel veel uit. Figuur 1 (zie p. 29) toont het gemiddelde verschil tussen de eerste en derde correctie per examen per school, uitgedrukt als een percentage van de maximaal te behalen score op het examen. Elke school uit het onderzoek is weergegeven met een cirkeltje. Hoe hoger het cirkeltje in de figuur, hoe groter het verschil in het voordeel van de eerste correctie, dat wil zeggen: hoe soepeler de eerste corrector in vergelijking met de derde correctoren.

Vierde correctie

Deze resultaten vereisten nader onderzoek. Om na te gaan of hier geen sprake was van een uitzondering is besloten de correctie opnieuw te laten doen en de uitkomsten daarvan te bespreken in een panel correctoren, het panelonderzoek vierde correctie (P4C). Hiermee kregen we een 'ware' beoordeling van de onderzochte examenwerken. Anders dan in het O2C-onderzoek ontvingen de pa- ►

► René Alberts en Bettina Ehrens, *Examenverslag 2012*, www.cito.nl.

► De andere vakken waren Nederlands en biologie vwo, Engels en wiskunde vmbo, tehatex en geschiedenis havo.

nakijken

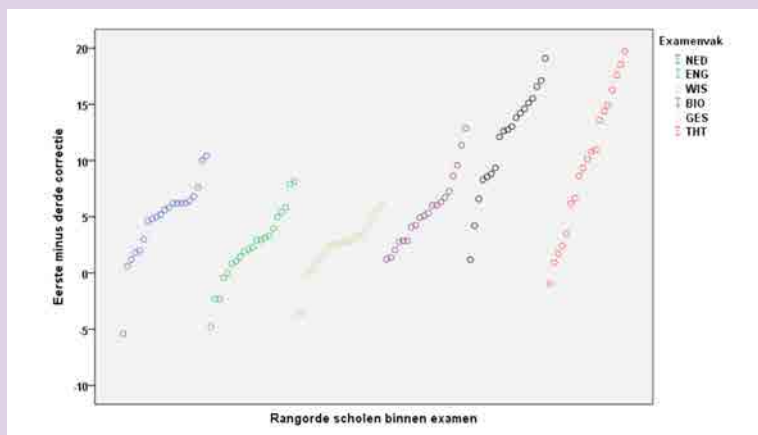


**'Door de derde correctie
gingen alle resultaten omlaag.'**

► Bij Nederlands is bij de panelbespreking veelvuldig gebruik gemaakt van de examenbespreking van Levende Talen. Als de aanvullingen op het cv van de regiobespreking van de VGN bij geschiedenis zouden zijn gebruikt, zou dit waarschijnlijk tot hogere scores hebben geleid.

► De vakken zijn verdeeld over de clusters talen, bèta, gamma en kunst. Geschiedenis was, samen met Duits, scheikunde en tehatex een van de vier havovakken.

nelleden een training in de interpretatie van het correctievoorschrift door de toetsdeskundige van Cito. Bovendien kregen zij volop de ruimte om met elkaar te discussiëren over de vraag, het beoordelingsmodel en de toegekende scores. Dit onderzoek werd uitgevoerd bij de vakken met de grootste verschillen: Nederlands, geschiedenis en tehatex. Per vak werden vijftien examenwerken opnieuw beoordeeld. Deze vijftien werken vormden geen willekeurige steekproef uit alle examenwerken van het hoofdonderzoek. Gekozen zijn de vijftien werken waarbij het verschil tussen de eerste en derde correctie het grootst was. De achterliggende gedachte was dat juist bij deze werken de kans op het vinden van te soepele of te strenge beoordelingen het grootst is. Naderhand zijn de scores van de eerste en derde correctoren vergeleken met die van het panel van vierde correctoren. Zo is geprobeerd te achterhalen in hoeverre de eerste correctoren



Figuur 1:
Gemiddeld verschil van de eerste minus de derde correctie per examen per school.

Te zien is dat de eerste correctie op vrijwel alle scholen hoger uitvalt dan de derde correctie. Gemiddeld over alle zes vakken gaat het om zes procent van de maximumscore. Daarnaast blijkt het verschil bij de examens geschiedenis en tehatex duidelijk groter dan bij de andere examens. Het grootste verschil zat bij geschiedenis, een heel punt. Als in percentages onvoldoenden gerekend zou worden, dan stijgt bij geschiedenis het aantal onvoldoenden van 23 procent naar 53 procent. Alleen tehatex kende een groter verschil (van 22 procent naar 54 procent). De andere vakken bleven hierbij achter, de stijging varieerde van 23 procent naar 29 procent bij Engels tot een verschuiving van 20 procent naar 31 procent bij Nederlands. Ook behoorlijk grote verschillen.

Figuur 1 laat duidelijk zien dat er een verschil is tussen 'soepele' en minder soepele scholen. Dat heeft tot gevolg dat een examenkandidaat op de ene school met hetzelfde vaardigheidsniveau tot een lagere beoordeling kan komen dan op een andere school. Uiteraard is dit een ongewenste situatie.

te soepel hadden beoordeeld of dat de derde correctoren te streng waren geweest. Bij geschiedenis bleek de gemiddelde score van het panel veel dichterbij de derde dan de eerste (en tweede) correctie te liggen. Kennelijk waren de eerste (en tweede) correctoren te soepel en de derde correctoren slechts in beperkte mate te streng. Uitgedrukt in cijfers op het CSE geschiedenis komt het verschil tussen de eerste, derde en vierde correctie overeen met de cijfers 7.1, 5.1 en 5.0. Bij tehatex vertonen de resultaten van het panel hetzelfde beeld (7.6, 5.7 en 5.3). Nederlands kende een kleiner verschil, 7.1, 5.7 en 6.7.³

De panelleden beoordeelden ook in hoeverre de eerste corrector het correctievoorschrift bij de vijftien examenwerken juist toepaste. Na elk beoordeeld antwoord maakten de panelleden een keuze uit vier mogelijkheden om het gedrag van de eerste corrector te beschrijven: de eerste corrector heeft het correctievoorschrift correct toegepast, er is twijfel of de eerste corrector te soepel beoordeelde, er is zeker te soepel beoordeeld en er is te streng beoordeeld. De resultaten zijn weergegeven in figuur 2.

Grote correctiedruk

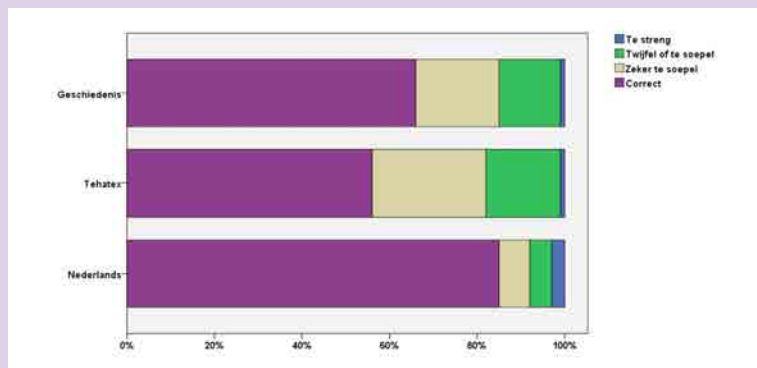
Het derde onderzoek is de enquête uit 2011-2012 naar de praktijk van de eerste en tweede correctie bij twaalf examenvakken verdeeld over vmbo-t, havo en vwo.⁴ Niet verrassend is dat geschiedenis het vak is met de grootste correctiedruk. Docenten havo geschiedenis besteden aan eerste en tweede correctie samen

63 uur, terwijl de gemiddelde docent (van alle docenten in het onderzoek) daaraan 35 uur besteedt. De verschillen tussen docenten zijn erg groot: de tien procent met de geringste correctielast besteedt er tien uur aan met ten hoogste 33 kandidaten, de tien procent met de zwaarste last zeventig uur voor 140 kandidaten. Op vrijwel alle scholen wordt noch de eerste, noch de tweede correctie gefaciliteerd. Dat leidt tot grote ontevredenheid over het schoolbeleid hierover bij bijna de helft van de ondervraagden.

De regels voor het uitvoeren van de tweede correctie zijn onlangs aangescherpt door OCW: de tweede correctie moet integraal (en niet steekproefsgewijs) worden uitgevoerd. Eén derde van de docenten houdt zich hieraan en voert ook de tweede correctie integraal uit. Geschiedenisdocenten houden zich vaak niet aan de aangescherpte regelgeving, omdat dit voor hen onuitvoerbaar is zonder facilitering door de school. De docenten geschiedenis geven aan dat zij twee keer zoveel tijd besteden aan de eerste als aan de tweede correctie. Meestal verloopt het overleg tussen eerste en tweede corrector soepel, telefonisch, en duurt het ongeveer vijftig minuten. Hulp van derden wordt zelden ingeroepen.

Aanbevelingen

De aanbevelingen die uit deze drie onderzoeken voortkomen, gaan in op verschillende onderdelen van het complexe geheel van de examencorrectie: de procedures en regelgeving, het schoolbeleid, deskundigheidsbevordering en de examenproductie. Deze aanbevelingen zullen nog bediscussieerd moeten worden door alle betrokkenen. Pas daarna kunnen de aanbevelingen uit de rapporten leiden tot een verbetering van de bestaande praktijk. De staatssecretaris heeft nu al een aantal conclusies getrokken. In zijn brief aan de Kamer bij de rapportage over dit onderzoek maakt hij bekend dat de Inspectie opdracht heeft gekregen bij de komende examens extra toe te zien op naleving van de eisen rond de tweede correctie.⁶ Er komen twee pilots waarin nader onderzoek wordt gedaan naar de examencorrectie. Daarnaast wil hij de schoolbesturen aanspreken op facilitering van de tweede correctie. Dit lijkt goed nieuws voor de geschiedenisdocenten maar zal voor de komende examens te laat komen. Gelukkig is er nu wel op alle niveaus aandacht voor de grote last van de examencorrectie bij geschiedenis. En dat



Figuur 2:
De mate waarin eerste correctoren het correctievoorschrift correct toepassen per examen.

Naar het oordeel van het panel geschiedenis heeft de eerste corrector het correctievoorschrift bij 66 procent van de nagekeken antwoorden correct toegepast. Bij veertien procent van de antwoorden stonden de vierde correctoren in dubio of de eerste corrector de kandidaat te soepel beoordeelde. Bij negentien procent van de antwoorden was het panel van mening dat de eerste corrector de kandidaat op een voor hen onverklaarbare wijze beoordeeld heeft. De eerste correctoren beoordeelden hier bijvoorbeeld volgens het principe 'alles wat niet echt fout is, is goed' of rekenden evident onjuiste antwoord goed. Daar staat tegenover dat de kandidaat bij één procent van de beoordelingen benadeeld is.

Daarnaast werd geconstateerd dat er eerste correctoren zijn die hun kandidaten 'onverklaarbaar toegeeflijk' beoordeelden. Die soepelheid kon niet worden toegeschreven aan een gebrek aan vakkenis, onvolkomenheden in het examen, de formulering van het antwoord door de kandidaat, etc.⁵

is goed nieuws voor de geschiedenisdocenten. Verbetering van de examencorrectie zal leiden tot een eerlijkere beoordeling van het werk van de kandidaten. En dat is goed nieuws voor elke betrokkene bij het centrale eindexamen. ■

Literatuur

- Hans Kuhlmeier en Ed Kremers, *De praktijk van de eerste en tweede correctie van het CSE, Verslag van een landelijke enquête*, Arnhem 2012;
- Hans Kuhlmeier en Ed Kremers, *De praktijk van de eerste en tweede correctie van het CSE, samenvatting van een onderzoek naar het functioneren van het CSE*, Arnhem 2012;
- Hans Kuhlmeier, Peter van Rijn en Ed Kremers, *Eerste, tweede en derde correctie van examenwerk: Wat is het verschil*, Arnhem 2012,
- Hans Kuhlmeier, Hugo Gitsels, Stefan Boom, Alex van de Kerkhof en Roel Sinkeldam, *Examenkenmerken en verschillen tussen correctoren, verslag van een panelonderzoek bij het CSE geschiedenis, tehatex en Nederlands*, Arnhem 2012.

► In H. Kuhlmeier, H. Gitsels, S. Boom, A. van de Kerkhof en R. Sinkeldam, *Examenkenmerken en verschillen tussen correctoren, verslag van een panelonderzoek bij het CSE geschiedenis, tehatex en Nederlands*, (Arnhem 2012) worden twee uitgewerkte voorbeelden gegeven van lastige vragen uit het onderzochte geschiedenisexamen (p. 14-17).
► <http://bit.ly/17g9069>